知识生产视角下学术论文质量自动评审指标体系构建研究*

■ 祝琳琳[1] 杜杏叶[1,2,3] 李贺[1]
[1] 吉林大学管理学院 长春 130022 [2] 中国科学院文献情报中心 北京 100190
[3] 中国科学院大学图书情报与档案管理系 北京 100190

摘要：[**目的/意义**]针对当前未发表学术论文质量的自动评审尚未形成统一的指标体系的问题，探索并建立一套具有引导性、科学性、客观性的论文质量自动评审指标体系，以提高评审效率。[**方法/过程**]在知识生产视角下，结合科学知识生产要素，分别从论文作者、参考文献、资金项目支持、选题、创新性、科学性、表达形式7个方面，构建论文质量自动评审指标体系，对其量化方法和技术进行简要说明，并运用主成分分析方法确定各项自动评审指标项权重及排序。[**结果/结论**]数据结果表明，论文科学性、创新性权重值均较高，论文表达形式中的摘要可读性、长度和参考文献所在期刊质量、新度同样是重要的自动评审因素，该结果能够为后续自动评审指标的量化处理提供借鉴。

关键词：知识生产 学术论文 质量 自动评审 主成分分析
分类号：G250
DOI：10.13266/j.issn.0252-3116.2018.24.011

1 引言

对于学术期刊，尤其是核心学术期刊来说，编辑部每年都会收到大量论文投稿，其工作量可想而知。通常每一篇学术论文稿件都会经过严格的评审过程，其中责任编辑初审、同行专家复审、主编终审的筛选过程更是会耗费大量人力物力和时间。最重要的是，其中涉及的人为干扰因素有很多，比如评价标准的主观性、专业认知的局限性等。稿源是学术期刊的核心，也是期刊的重要资源[1]。高质量的稿源带动期刊的发展，高质量期刊吸引优质的稿源，这样才能逐渐形成良性循环。

目前，对学术论文的普通评价方法主要分为定性评价与定量评价。定性评价包括同行评议，即指专家根据自身经验与知识进行主观评判；定量评价包括引文分析评价，指根据论文发表后的引用次数评价其学术影响力，另外，也有学者基于社交媒体评价的替代计量学方法[2]、PageRank改进方法[3]等定量评价方法对论文进行评价。现实情况下，编辑部收到的学术论文投稿均是未发表的论文，缺少引用量、期刊影响因子等外部指标，而且无论是定性评价还是定量评价方法，普通评价方法都是依据个人主观评价和面向已发表论文的评价。因此，区别于论文的定量评价，本文中论文质量自动评审的对象为投稿中且未发表的论文，即没有任何外部指标，仅根据学术论文中所涵盖的内部知识单元结构和内容，对其进行综合评价；区别于专家的人工定性评价，论文质量的自动评审需要运用知识挖掘、主题分析等技术辅助得以自动实现。

关于未发表学术论文质量的自动评审尚未形成统一的指标体系，因此，本文拟通过知识生产视角，结合知识生产要素构建学术论文质量自动评审指标体系，并运用SPSS和Excel软件，以主成分分析方法确定多指标项权重。本文将探索并建立一套具有引导性、科学性、客观性的论文质量自动评审指标体系，从而为后续自动评审指标的量化处理提供参考和基础。

2 学术论文质量自动评审指标体系构建

2.1 相关理论——知识生产(knowledge production)

2.1.1 知识生产内涵 1996年，国际经济发展与合作组织(OECD)发表了《以知识为基础的经济》报告，认

* 本文系中国科学院文献情报能力建设项目“新型出版研究”子项目“稿件自动评审研究”(项目编号：院1732)研究成果之一。
作者简介：祝琳琳(ORCID:0000-0003-3749-6954)，博士研究生；杜杏叶(ORCID:0000-0001-5016-0561)，副研究馆员，副编审，博士研究生，通讯作者，E-mail:duxy@mail.las.ac.cn；李贺(ORCID:0000-0001-8847-3619)，教授，博士生导师。
收稿日期:2018-06-07 修回日期:2018-09-20 本文起止页码:79-86 本文责任编辑:易飞

为知识生产就是开发与提供新知识,是探索性和创造性的智力活动。知识生产并没有统一的定义,但是都体现出了其利用知识创造新知识的“知识再生产”观点,即知识生产是人们通过脑力劳动创造出新知识的过程,是在已有知识的基础上发现新知识的过程[4]。知识生产的实质就是知识创新,就是指科学发现和技术发明,这两类发现和发明可通过不同的成果形式存在,包括科学论文、著作、报告、图纸、电脑程序等[5]。基于宏观经济的角度,何传启等[6]认为知识生产是一种生产活动, 并指出知识生产是从事研发活动的科学家、工程师、博士研究生及助手等人员所进行的知识创新。

通常,知识生产可以划分为 3 种类型,即原创性知识生产、复制性知识生产、定制性知识生产。科学知识生产主要是原创性的知识生产,学术论文的产出即是科学知识生产过程下的成果。科学知识生产首先要从选题开始,然后进行信息收集和研究工作。选题、信息收集和信息分析过程是认识不断深化和逐步提高的过程;分析研究是知识生产过程中非常重要的环节,如定量分析、定性分析;撰写论文是知识生产的主要工序,也是最重要的环节,完成此环节后,知识生产过程结束,形成完整的知识产品[7],如图 1 所示:

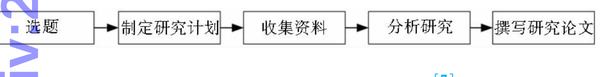


图 1　科学知识生产一般步骤[7]

2.1.2　知识生产要素　知识生产包括两个途径,即组合和交换。将获取的原料、资源等生产要素进行重新组合以产生新的生产方式,而当有限的资源被不同的行为主体拥有的时候,资源的相互交换就成了资源组合的先决条件[7]。通过知识生产的途径可知,知识生产需要一定的知识资源基础,同时需要具有一定信息素养能力的知识工作主体来完成。焦桐等在研究中指出,实现知识生产需要具备 4 个要素,即直接或间接需要的有关知识、作为主体的智力劳动者、生产知识的工具、相应的组织和管理[8]。可以说,科学知识生产就是科研主体在科研资源条件的支持下,通过一定的生产关系,作用于科研知识基础,创造出科研成果[9]。

李正风在其研究中指出,科学知识生产所需要的、可支配、可利用的要素总和即是资源禀赋,表征着科学知识生产方式的基础与条件[10]。其中与学术论文知识生产相关的资源包括人力资源、财力资源、物力资源、信息资源。根据文献[10]的解释说明,人力资源是指从事科学知识生产的人员的数量与质量,包括是否具有高水平科学家、高层级学术职衔等;财力资源是指可用于科学知识生产的资金;物力资源是指支撑科学知识生产活动的科学设备、基础设施和相关条件平台;信息资源是指人类科技活动所产生的基本科学技术数据、资料,以及面向不同需求加工整理形成的各种科学数据产品和各种载体的科学知识产品,通常包含在科学图书、期刊、报告、论文、档案等科学文献中。

2.2　学术论文质量自动评审指标筛选与确定

一篇学术论文从选题构思开始,直至撰写完毕,就是一个科学知识生产的标准过程。同样,知识生产过程中涉及的生产要素同样适用于学术论文的完成。为了尽可能地全面考量学术论文质量评价指标,本文基于知识生产视角,结合学术论文内部知识单元结构及内容,总结知识生产要素,构建学术论文质量自动评审指标体系,即包括科研主体及其生产关系、科研知识基础、科研资源条件和科研成果本身 4 个方面[9]。

2.2.1　科研主体及其生产关系　学术论文的撰写也是论文创建的过程,它的质量取决于撰写者(作者)的水平和努力程度。这个过程受到个人能力、知识水平和努力程度等因素的影响[11]。因此,在对一篇学术论文质量进行评价时,应该对其科研主体及其生产关系,即作者进行分析。

蓝华等[1]提出影响科技期刊质量的 4 个因素,其中就包含对稿源质量的评价,并从作者评价这一角度,对作者学术素养、作者机构分布进行了评价。其中作者学术素养包括对其职称学历的参考,如作者中高级职称的比例、博士的比例、博士研究生的比例等;作者机构分布包括对其来源单位的参考,如“985”“211”高校比例、其他科研机构比例等。同时,作者声誉、科研能力也应是评价中的重要参考,虽然单篇学术论文稿件没有引用量,但可以对其作者历史发文数量以及历史发文引文量进行分析[12],另外也可以根据 J. E. Hirsch 提出的 H 指数对科研人员的学术产出进行评价[13]。

现如今科研论文的多作者合作变得非常普遍,在以被引次数高低等价于学术论文影响力高低的基础下[14],F. Didegah[15]在研究中指出,与单一作者相比,多作者合作的文章会得到更多的引用量,并利用相关和回归测试说明作者数量以及不同机构研究人员的机构合作均与较高的引用相关。R. Sooryamoorthy[16]也认为,国际合作会导致论文引用次数的增加。林德明等[17]根据对科技论文评价体系的研究,提出一种对作者合作规模的测度公式,即作者合作规模 = 合作学科

数×地区数×作者数/学科平均作者数。

因此,结合知识生产下的科研主体及其生产关系这一要素,本文将该要素下论文质量自动评审的指标划分为作者学术素养、作者科研能力、作者声誉、作者数量、作者来源机构、作者跨国家/地区合作、作者跨学科合作。为了实现论文质量自动评审,需要对各项指标进行量化处理,详细指标及参考量化方法如表1所示:

表1　科研主体及其生产关系要素下论文质量自动评审指标及量化方法

论文质量自动评审指标	辅助实现指标量化的方式/技术方法参考
作者学术素养	作者中各职称、各学历的比例
作者科研能力	作者H指数、历史发文数、历史发文引文数
作者声誉	作者专家称号,按照专家称号级别划分赋值
作者数量	作者总人数
作者来源机构	作者来源单位数量
作者跨国家/地区合作	作者所在国家/地区数量
作者跨学科合作	作者中所涵盖的学科数量

2.2.2　科研知识基础　参考文献是学术论文中的重要组成部分,并在一定程度上影响论文的质量。在知识生产视角下,参考文献作为生产要素中的信息资源,其数量与质量更是反映了论文作者对于过去相关研究的了解程度与吸收程度。因此,应对论文的科研知识基础,即参考文献进行分析。

国内外研究学者研究认为,参考文献数量与被引频次有相关性[18-19],同时研究表明,论文质量除了与本身质量相关联外,引用了高影响力参考文献的文章被引用的次数更多,而且参考文献的平均引用次数、引文总数和参考文献的h指数相关[20]。在衡量参考文献的质量方面,林德明等[17]通过参考文献篇均被引频次和参考文献的所在期刊的篇均影响因子对其进行测度。参考文献的引用半衰期与被引频次呈现显著负相关,即参考文献的新度与文章被引频次相关[9]。

因此,结合知识生产下的科研知识基础这一要素,本文将该要素下论文质量自动评审的指标划分为参考文献数量、参考文献被引频次、参考文献所在期刊质量、参考文献新度。详细指标及参考量化方法如表2所示:

表2　科研知识基础要素下论文质量自动评审指标及量化方法

论文质量自动评审指标	辅助实现指标量化的方式/技术方法参考
参考文献数量	参考文献总数
参考文献被引频次	参考文献总被引频次
参考文献所在期刊质量	参考文献所在期刊影响因子总和
参考文献新度	参考文献引用半衰期

2.2.3　科研资源条件　学术论文是否由资金项目资助能够在一定程度上体现出该科研项目论文成果的质量:①科研项目立项已经经过专家评议,说明该课题研究的价值;②基金项目资助的项目成果具备一定的财力支撑,财力资源可用于科学论文的知识生产过程;③在财力资源的支撑下,有利于科学研究所需设备、基础设施、相关条件的提升。因此,应该对论文稿件的科研资源条件,即资金项目支持情况进行分析。

张秀华等[21]指出,项目资助资金对现代知识生产机制的影响越来越大,在学界被认为级别比较高的基金有国家自然科学基金(NSFC)、国家社会科学基金、"863"项目资金等,已成为衡量一个学校或者学术研究团体科研能力的重要评价指标。张诗乐等[22]根据国家自然科学资金资助项目和SCI论文的数据统计,说明我国NSFC支持下的SCI论文整体学术水平和科研影响力相对较高。

因此,结合知识生产下的科研资源条件这一要素,本文将该要素下论文质量自动评审的指标划分为是否有资金项目支持和资金项目支持级别。详细指标及参考量化方法如表3所示:

表3　科研资源条件要素下论文质量自动评审指标及量化方法

论文质量自动评审指标	辅助实现指标量化的方式/技术方法参考
是否有资金项目支持	有无资金项目支持,可设为二分类变量
资金项目支持级别	按照国家级、省级、市级、校级划分赋值

2.2.4　科研成果本身　一篇完整的学术论文包括题目、作者、摘要、关键词、正文、参考文献等众多知识单元组成结构。论文作为知识生产和创新的成果,要通过知识安排、知识编辑和整理各种信息来源完成知识内化,使内部知识有序化[11],因此学术论文本身的质量至关重要,其必须具有科学性、紧密性和逻辑性。

有学者基于国内核心期刊论文评审表选用的评议指标进行了统计分析,认为在论文评审中应注重对选题、创新性、科学性、写作水平以及研究价值的评价[23]。闻浩等[24]认为,科技论文的评审应从内容质量和形式质量两个部分进行。其中内容质量审定主要从科学性、创新性和知识性等方面开展,形式质量审定要通过结构架构、行文格式和表述形式等方面开展。钟灿涛等[25]借鉴国际标准化组织的质量概念定义,认为应从学术论文内容和表达形式(可读性)满足学术交流、知识积累以及其他相关方要求程度的方面进行评价。学术论文应具有创新性、准确性,而且应做到格式标准、表达准确。其中,创新性应做到选题新颖、研究

方法新颖，具有一定前瞻性；准确性应做到数据准确无误、逻辑性强、层次清晰、论据充分；格式标准应做到论文撰写符合国家规范或标准、术语准确、图表清晰；表达准确应做到学术水平高并且表述流畅[26]。同时，笔者查阅了全国优秀博士学位论文评选标准以及高等学校科学研究优秀成果奖（人文社会科学）的评定办法[27-28]，二者均对研究成果创新性和选题前沿性进行了明确规定。

因此，科研成果本身应作为学术论文质量自动评审指标中重点考查要素，笔者结合各研究结论以及全国优秀博士论文等评价标准，从论文选题、创新性、科学性、表达形式 4 个方面提出评价指标，具体设定如下：

（1）论文选题。论文选题通常是指论文主题所论述或研究的主要内容，主题应具有新颖度与热度，这是保证学术论文质量的因素之一。主题新颖是学术创新评价最本质的特征之一，是指选题主题的研究方向为学科前沿，具有开创性。主题新颖度是可以量化测算的，如杨建林等[29]利用 KPTIDF 方法对主题新颖度进行量化，逯万辉等[30]基于 Doc2Vec 和 HMM 算法测度学术成果的主题新颖度；主题热度体现当前科学前沿主题时效性高，如冯佳等[31]利用 LDA 主题模型抽取科学前沿研究主题。

（2）论文创新性。论文创新性主要在于新论点和新论据，新论点包括新理论、新方法、新对策和新学科，新论据包括新数据和新事实[32]。学术论文成果应该在理论、技术或研究方法等微观方面体现出一定的创新性。对于创新性的测度，可以基于知识元理论，利用知识元抽取技术抽取论文文本中的理论、技术、方法知识元，分别构建理论知识库、技术知识库、方法知识库，通过知识元标注、知识元链接、知识元集成等相关技术，将评审论文中的知识元与知识库进行关联、对比和计算，实现理论、技术、方法创新性的自动评审过程[33]。

（3）论文科学性。学术论文应具有科学性，具体体现在其技术水平、研究方法、实验设计、数据处理等方面。在技术水平测度方面，可以借鉴技术进化树概念，依照时间轨迹，探究技术演化路径。另外，同样借鉴论文创新性研究的知识元理论，可以构建研究方法、实验设计、数据处理知识库，比如通过数据处理方法与数据处理规则的映射，对其数据处理的科学性进行自动判断。

（4）论文表达形式。学术论文的表述应做到清晰明了。主要包括摘要长度、关键词数、标题长度、篇幅、中英文摘要可读性、图表清晰度。其中中英文摘要文字的可读性，可利用中文 Flesch 易读度和英文可读性系数，如 Kincaid、ARI、Flesch score、Fog index 等进行自动评判[34]；图表清晰度可利用图像识别技术进行判定。

此处指标的处理需要应用大量自动处理算法及技术，详细指标及参考量化方法如表 4 所示：

表 4　科研成果本身要素下论文质量自动评审指标及量化方法

论文质量自动评审指标	辅助实现指标量化的方式/技术方法参考
论文选题新颖度	文本主题新颖度测度方法，如 KPTIDF 等
论文选题热度	文本主题热度测度方法，如 LDA 等
论文理论创新性	基于知识元理论，利用大量论文数据构建理论知识库、技术知识库、方法知识库，通过抽取、标注、关联、计算等处理过程进行赋值
论文技术创新性	
论文研究方法创新性	
论文技术水平科学性	技术进化树
论文研究方法科学性	基于知识元理论，利用大量论文数据构建方法知识库、实验设计知识库、数据处理知识库，通过抽取、标注、计算、映射等处理过程进行赋值
论文实验设计科学性	
论文数据处理科学性	
论文摘要长度	摘要字符数
论文关键词数	关键词数
论文标题长度	标题字符数
论文篇幅	正文字符数
论文中英文摘要可读性	英文可读性系数、中文 Flesch 易读度等
论文图表清晰度	图像识别技术等

因此，基于上述 2.2.1－2.2.4 节的分析，建立学术论文质量自动评审指标体系，见图 2。

3　学术论文质量自动评审指标权重分析

目前，多因素综合评价指标权重确定的方法包括主观赋权法与客观赋权法，本文应用客观赋权法中的主成分分析方法确定多指标权重。该方法通过线性变换，用较少指标表达大部分信息量，在无需指标样本数据的情况下利用主成分分析方法基本原理，建立权重模型，从理论证明到实际数据应用具有可行性、有效性和通用性。

3.1　计算步骤

（1）确定指标因素后，由具有相关领域知识的专家对各个指标进行评分。评价等级分为 5、4、3、2、1，即非常重要、比较重要、一般重要、不太重要和不重要，通过五级量表对各指标因素的重要程度进行评分。

（2）主成分分析通常是将原来的指标通过线性组合重新组成彼此无关的综合指标，即主成分线性组合模型。表达式如公式（1）所示：

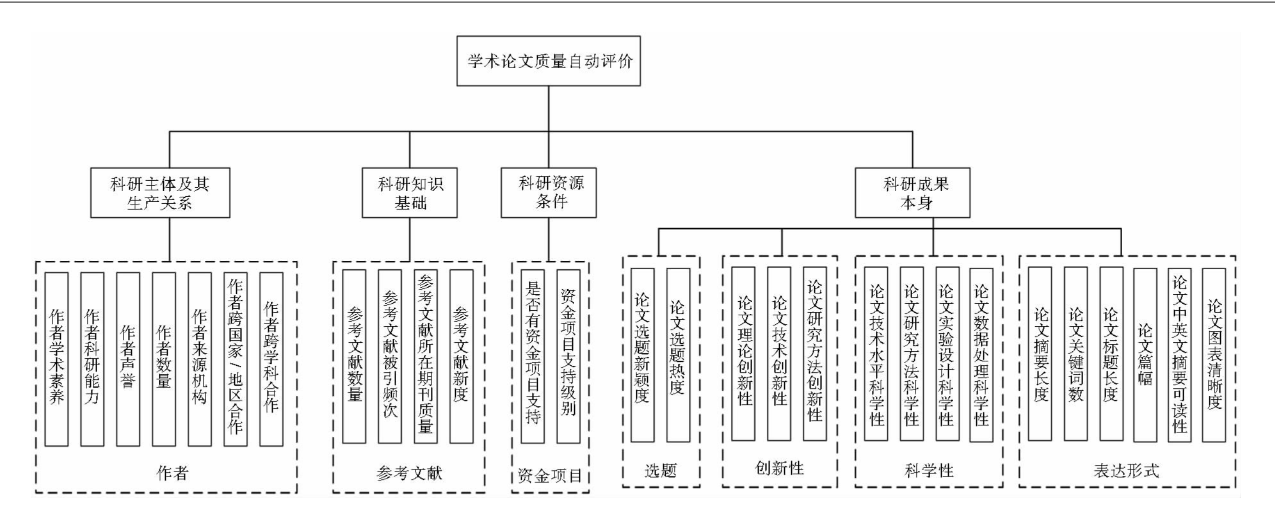


图2　学术论文质量自动评审指标体系

$$\begin{cases} F_1 = u_{11}x_1 + u_{21}x_2 + \cdots + u_{h1}x_h \\ F_2 = u_{12}x_1 + u_{22}x_2 + \cdots + u_{h2}x_h \\ \vdots \\ F_m = u_{1m}x_1 + u_{2m}x_2 + \cdots + u_{hm}x_h \end{cases} \quad 式(1)$$

$$u_{ij} = \frac{f_{ij}}{\sqrt{\lambda_j}}, j = 1,2,\cdots,m \quad 式(2)$$

其中，h 为评价指标的个数，x 为评价指标，该部分可经过 SPSS 软件的主成分分析得出，在公式(1)中，$F_1, F_2, \cdots, F_m$ 是 m 个主成分；u_{ij}是决策矩阵系数，需要经过公式(2)中的计算方法，利用软件得出的初始因子载荷 f_{ij} 和特征根 λ_j 进行转化。

(3)对指标在 m 个主成分线性组合中的决策矩阵系数 u_{ij} 做加权平均，计算各指标权重 ω_i，如公式(3)所示，其中，V 为各主成分方差贡献率。

$$\omega_i = \frac{u_{i1}V_1 + u_{i2}V_2 + \cdots + u_{im}V_m}{\sum_{n=1}^{m} V_n}, i = 1,2,\cdots,h \quad 式(3)$$

(4)若生成负权数，可将 ω_i 通过公式(4)和公式(5)中 k 的取值规则转化成 ω_i'，即通过平移使之归结为结构相对数，然后将 ω_i'归一化[35]；若未生成负权数，则直接将 ω_i 归一化。

$$\omega_i' = \frac{k + \omega_i}{2k} \quad 式(4)$$

$$k > \max_{1 \le i \le h} |\omega_i| \quad 式(5)$$

因此，根据步骤(1)－(4)，即可通过主成分分析方法得出各指标的权重 $W = (\omega_1, \omega_2, \cdots, \omega_h)$。

3.2　实证研究结果

3.2.1　描述性统计结果　本文通过问卷调查的方式，主要对图书情报与档案管理专业学者和该专业期刊编辑部工作人员进行调查。被调查者根据自身学术论文评审经历和投稿经历，对 28 个指标因素相对于学术论文稿件质量自动评审影响的重要程度进行评分。该问卷发放时间为 2018 年 5 月 9 日至 2018 年 5 月 17 日，共收集问卷 209 份，删除填写时间过短或过长的问卷，共得到有效问卷 180 份。本问卷调查范围涉及多个地区，分布情况见图 3。由于被调查者可能兼具高校教师和编辑部主编等多重身份，各身份数量之和大于有效问卷数目，详细身份分布见图 4 。

3.2.2　权重确定结果　本文结合 SPSS 和 EXCEL 软件，对数据进行处理。在计算权重之前，需要对数据的信度和效度进行检验，该问卷全部指标题项的 Cronbach α 系数为 0.883，说明问卷的整体信度较高。在对各个维度的分量表进行信度检验时，剔除了校正的项总计相关性(CITC)低于 0.5 而且项已删除的 Cronbach α 系数高于各维度分量 Cronbach α 系数的指标，包括“作者学术素养”“作者科研能力”“作者声誉”“作者数量”“论文图表清晰度”5 项指标，各维度分量 Cronbach α 系数均高于 0.7。另外，通过 KMO 和 Bartlett 的检验，KMO 值为 0.802，显著性为 0.000，说明量表具有良好的建构效度。

根据特征值大于 1 的原则，确定主成分个数为 6，用抽取出的 6 个主成分代替原有的多个指标。前 6 个特征根的可解释总体方差为 69.935%，详见表 5 解释总方差。根据 3.1 节介绍的计算步骤，结合成分矩阵、特征值与方差贡献率数据，计算出本文各指标因素权重，见表 6，其中加粗数值代表该权重值大于平均权重值 0.0435(1/23)。

3.3　结果分析与讨论

(1)结果显示，学术论文科学性的 4 项指标与学术论文创新性的 3 项指标权重值均位于前列，尤其是研

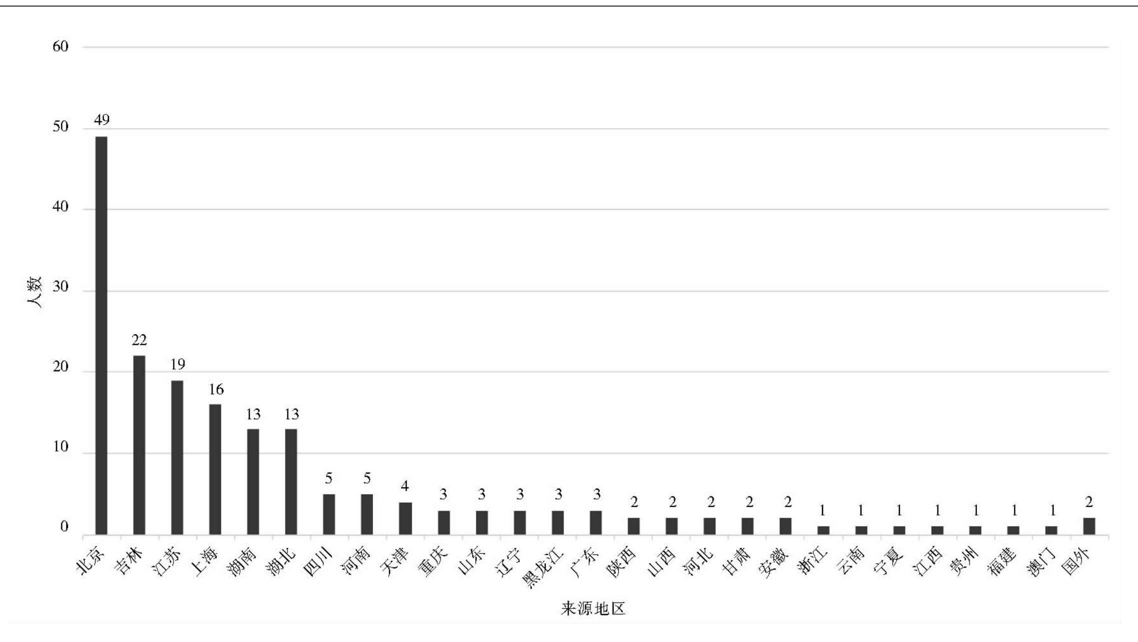


图 3 被调查者来源地区分布情况

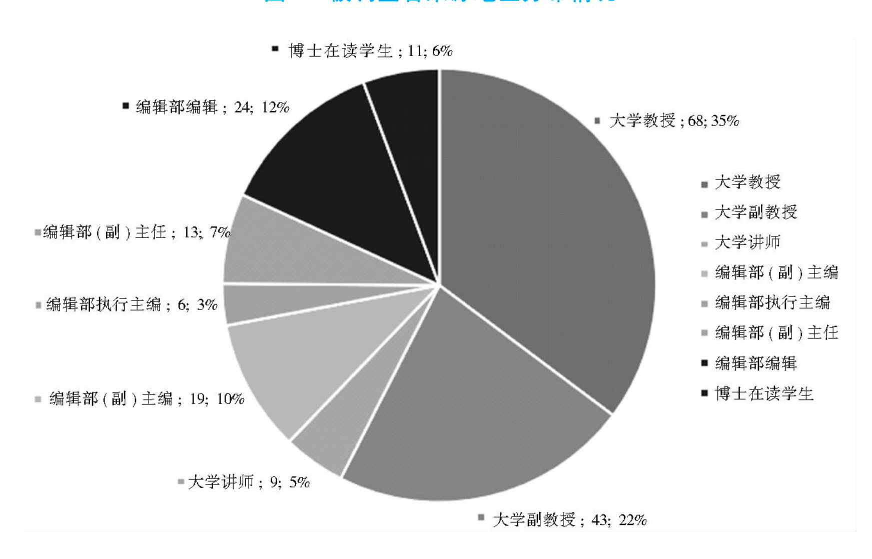


图 4 被调查者身份分布

表 5 解释总方差

成份	初始特征值		
	特征根	方差的 %	累积 %
1	6.244	27.148	27.148
2	3.886	16.897	44.046
3	1.953	8.490	52.536
4	1.708	7.425	59.960
5	1.243	5.403	65.364
6	1.051	4.571	69.935

表 6 各指标因素权重值

序号	指标	权重	序号	指标	权重
1	论文实验设计科学性	0.057	12	论文关键词数	0.042
2	论文数据处理科学性	0.057	13	论文篇幅	0.042
3	论文研究方法科学性	0.056	14	论文标题长度	0.040
4	论文技术水平科学性	0.056	15	作者跨国家地区合作	0.040
5	论文技术创新性	0.049	16	参考文献被引频次	0.039
6	论文研究方法创新性	0.049	17	参考文献数量	0.038
7	论文中英文摘要可读性	0.049	18	作者来源机构	0.037
8	参考文献所在期刊质量	0.045	19	作者跨学科合作	0.037
9	论文摘要长度	0.045	20	论文选题新颖度	0.036
10	论文理论创新性	0.044	21	是否有资金项目支持	0.035
11	参考文献新度	0.044	22	资金项目支持级别	0.034
			23	论文选题热度	0.029

究方法、数据处理、实验设计等的科学性权重更是达到 0.057 和 0.056，可见论文本身知识结构的科学性与创新性对于其质量自动评审有着非常重要的作用。

（2）结果中关于摘要可读性和长度的指标权重达到了 0.049 和 0.045，可见学术论文的表达形式同样是

重要的自动评审因素。摘要是对文章的凝练,因此无论在论文质量评审还是在读者阅读文献的过程中,摘要都是论文的重要组成部分,应该做到表述精准、说明详尽、概括简练,使广大读者能够准确了解其中的思想内涵。另外,一些看似与论文内容无关的因素,如关键词数、篇幅、标题的权重值均达到0.04以上,也应该在论文质量自动评审中得到一定的重视。

(3)结果显示,参考文献所在期刊质量、参考文献新度指标的权重值达到0.045和0.044,而参考文献被引频次和数量的权重则相对较低。结果说明,在科研知识基础维度,评审中更加关注参考文献的“质”,而不是“量”,论文参考文献所在期刊质量和新度对于其质量自动评审有着非常重要的作用。

(4)结果中关于学术论文作者方面的指标权重略低。在论文的初审阶段,可以对作者信息进行一定的评判,避免误审等情况的发生,但并不是主要的评审要素。

(5)在学术论文选题方面和资金项目支持方面,各指标的权重结果相对较低。结果说明,被调查对象对于选题新度、热度与资金项目支持并不是十分关注,研究主题的新颖度并不是决定学术质量的绝对因素[30]。因此,在论文的自动评审中虽然需要对二者进行考量,但并不是主要的评审要素。

综上所述,每一个指标均在不同程度上对学术论文质量自动评审存在着一定的影响。但是,其中几个权重值较高的因素是影响学术论文质量自动评审的关键因素,且大多集中在科研成果本身这一内部结构上,分别为科学性(论文实验设计科学性、数据处理科学性、研究方法科学性、技术水平科学性)、创新性(论文技术创新性、研究方法创新性、理论创新性)、表达形式(论文中英文摘要可读性、摘要长度)和参考文献(所在期刊质量、参考文献新度)。

4 结语

本文基于知识生产视角,结合科学知识生产要素,通过文献调研与梳理,从科研主体及其生产关系、科研知识基础、科研资源条件和科研成果本身4个维度出发,分别从论文作者、参考文献、资金项目支持、选题、创新性、科学性、表达形式7个角度,构建学术论文质量自动评审指标体系。本文一方面对各个因素指标项的具体量化方法和技术进行了简要说明;另一方面,通过问卷调查获取专家数据,运用主成分分析方法确定各项质量自动评审指标项的权重,并根据权重值对各因素指标项进行排序。

本文调查的专家专业多为图书情报领域,可能在学科方面存在一定局限性。在后续研究中,将在各项自动评审指标项权重的基础上,通过各种技术手段,对各项指标进行实际量化和计算,进一步实现学术论文自动评审过程。

参考文献:

[1] 蓝华,刘微微,于渤.基于影响要素的科技学术期刊质量模糊综合评价[J].中国软科学,2009(1):144－150.

[2] PRIEM J,PIWOWAR H A,HEMMINGER B M. Altmetrics in the wild: using social media to explore scholarly impact[EB/OL]. [2018－04－10]. https://arxiv.org/abs/1203.4745.

[3] 郑美莺,梁飞豹,梁嘉熹.单篇论文评价方法——PaperRank算法[J].科技与出版,2016(7):94－98.

[4] 李进才.高等教育世纪观[M].武汉:武汉出版社,2001.

[5] 袁志刚.论知识的生产和消费[J].经济研究,1999(6):59－65.

[6] 何传启,张凤.知识创新[M].北京:经济管理出版社,2001.

[7] 杨丹.网络时代的社会科学知识生产[M].北京:社会科学文献出版社,2009.

[8] 焦桐,肖源.大数据搜索引擎下的知识产出机制研究[J].情报科学,2018, 36(3):33－38.

[9] 徐庆富,康旭东,张春博.多期刊比较视角下的论文被引频次若干影响因素研究[J].情报杂志, 2018(2):147－153.

[10] 李正风.科学知识生产方式及其演变[M].北京:清华大学出版社,2006.

[11] LIU G L,WANG D. The Study on double-element control mechanism of academic quality of sci-tech papers published via Internet [C]//IEEE. International conference on multimedia and information technology. China:IEEE Xplore Digital Library, 2009:681－684.

[12] WANG M,YU G,AN S,et al. Discovery of factors influencing citation impact based on a soft fuzzy rough set model[J]. Scientometrics,2012,93(3):635－644.

[13] HIRSCH J E. An index to quantify an individual's scientific research output[J]. Proceedings of the National Academy of Sciences of the United States of America,2005,102(46):16569－16572.

[14] GARFIELD E. Citation indexes for science: a new dimension in documentation through association of ideas[J]. Science,1955,122(3159):1127－1128.

[15] DIDEGAH F. Investigating different types of research collaboration and citation impact: a case study of Harvard University's publications[J]. Scientometrics,2011,87(2):251－265.

[16] SOORYAMOORTHY R. Do types of collaboration change citation? collaboration and citation patterns of South African science publications[J]. Scientometrics,2009,81(1):177－193.

[17] 林德明,姜磊.科技论文评价体系研究[J].科学学与科学技术管理,2012,33(10):11－17.

[18] GREGORY D W, PETER K J, TATIANA O S. Hot topics and popular papers in evolutionary psychology: analyses of title words and citation counts in evolution and human behavior, 1979 – 2008[J]. Evolutionary psychology, 2009, 7(3): 348 – 362.
[19] 简琳, 何静, 周剑. 论文被引的文本因素分析: 多学科视角[J]. 图书情报工作, 2011, 55(20): 32 – 35.
[20] BORNMANN L, SCHIER H, MARX W, et al. What factors determine citation counts of publications in chemistry besides their quality? [J]. Journal of informetrics, 2012, 6(1): 11 – 18.
[21] 张秀华, 靳茜. 从国家社科基金项目看图书情报知识生产[J]. 大学图书馆学报, 2012, 30(1): 17 – 22.
[22] 张诗乐, 盖双双, 刘雪立. 国家自然科学基金资助的效果——基于论文产出的文献计量学评价[J]. 科学学研究, 2015, 33(4): 507 – 515.
[23] 廖建桥, 文鹏, 胡凌芳. 我国学术论文评审标准研究[J]. 科学学研究, 2010(8): 1128 – 1134.
[24] 闻浩, 鲁立. 科技期刊论文质量评价方法及相关研究[J]. 情报工程, 2017, 3(5): 60 – 67.
[25] 钟灿涛, 万猛, 王萍. 网络时代学术论文质量控制和标记系统案例研究[J]. 情报科学, 2012, 30(6): 893 – 898.
[26] 何丽. 科技期刊论文评审机制研究[J]. 科技创业月刊, 2014 (4): 158 – 160.
[27] 中华人民共和国教育部. 关于印发《全国优秀博士学位论文评选办法》的通知[EB/OL]. [2018 – 04 – 02]. http://old.moe.gov.cn//publicfiles/business/htmlfiles/moe/moe_841/201002/xxgk_82676.html.
[28] 中华人民共和国中央人民政府. 教育部关于印发《高等学校科学研究优秀成果奖(人文社会科学)奖励办法》的通知[EB/OL]. [2018 – 04 – 02]. http://www.gov.cn/gongbao/content/2009/content_1388675.htm.
[29] 杨建林, 钱玲飞. 基于关键词对逆文档频率的主题新颖度度量方法[J]. 情报理论与实践, 2013, 36(3): 99 – 102.
[30] 逯万辉, 谭宗颖. 学术成果主题新颖性测度方法研究——基于 Doc2Vec 和 HMM 算法[J]. 数据分析与知识发现, 2018(3): 22 – 29.
[31] 冯佳, 张云秋. 基于 LDA 和本体的科学前沿识别与分析方法研究[J]. 情报理论与实践, 2017, 40(8): 49 – 54.
[32] 周露阳. 论审评学术论文创新因素的指标体系[J]. 编辑学报, 2006 (1): 68 – 70.
[33] 刘杰, 秦春秀, 赵捧未, 等. 基于知识元的科技文本资源内容组织方法[J]. 情报理论与实践, 2018, 41(4): 128 – 133.
[34] DIDEGAH F, THELWALL M. Which factors help authors produce the highest impact research? collaboration, journal and document properties[J]. Journal of informetrics, 2013, 7(4): 861 – 873.
[35] 张鹏. 基于主成分分析的综合评价研究[D]. 南京: 南京理工大学, 2004.

作者贡献说明:

祝琳琳: 提出研究思路、研究方法, 撰写论文初稿;

杜杏叶: 负责数据收集与分析, 参与研究思路讨论, 修改论文;

李贺: 负责论文写作指导与审阅, 最终定稿。

Study on the Construction of Index System for Automatic Review of Academic Paper Quality Under the Perspective of Knowledge Production

Zhu Linlin[1] Du Xingye[1,2,3] Li He[1]
[1] School of Management, Jilin University, Changchun 130022
[2] National Science Library, Chinese Academy of Sciences, Beijing 100190
[3] Department of Library, Information and Archives Management, University of Chinese Academy of Sciences, Beijing 100190

Abstract: [**Purpose/significance**] The automatic review of the quality of unpublished academic papers has not yet formed a unified index system, so this article explores and establishes a set of leading, scientific and objective index system for automatic review to improve the efficiency. [**Method/process**] From the perspective of knowledge production, combined with the factors of scientific knowledge production, the index system of academic paper quality for automatic review whose quantitative methods and techniques are briefly introduced, is constructed from seven aspects of author, reference, fund project support, selection of topics, innovation, scientificity and expression form. The principal component analysis method is used to determine the weight and the ranking of various automatic review indexes. [**Result/conclusion**] The results show that the weight value of the scientificity and innovation of a paper are high. The readability, length of the abstract and the quality and newness of the journal in which the references are published are also important factors. The results can provide references for the quantitative treatment of the follow-up automatic review indexes.

Keywords: knowledge production academic papers quality automatic review principal component analysis